DOCUMENT RESUME

ED 295 492 FL 017 441

AUTHOR Wang, Lih Shing; Stansfield, Charles W.
TITLE Chinese Proficiency Test. Test Interpretation Manual.
INSTITUTION Center for Applied Linguistics, Washington, D.C.
PUB DATE 88
NOTE 14p.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Chinese; Difficulty Level; Item Analysis; Language Proficiency; *Language Tests; Listening Comprehension; Oral Language; Reading Comprehension; Scoring; *Test Format; *Test Interpretation; Test Manuals; Test Reliability; *Test Use; Test Validity; Uncommonly Taught Languages; Written Language
IDENTIFIERS *Test Content

ABSTRACT

The manual for administration of the Chinese Proficiency Test contains an overview of the program, including: (1) its history, content, and format; (2) its primary focus and uses; (3) administration procedures, including registration, ordering the test, reporting scores, and billing; (4) the interpretation of test scores based on normative data provided in the manual; and (5) the test's statistical characteristics, including reliability, standard error of measurement, content and concurrent validity, difficulty based on item analysis, and intercorrelation among subtest scores. An ordering form for the test is appended. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED295492

漢 語 能 力 測 驗

CHINESE PROFICIENCY TEST

TEST INTERPRETATION MANUAL

By

Lih Shing Wang

Charles W. Stansfield

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.R. Tucker
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

This manual is designed for individuals who use or interpret CPT scores. It presents an overview of the CPT program, uses of the CPT, test administration procedures, interpretation of test scores, and statistical characteristics of the test.

Center for Applied Linguistics
1118 22nd Street, N.W.
Washington, DC 20037
(202) 429-9292

Copyright (c) 1988 by Center for Applied Linguistics, Washington, DC.
All rights reserved.

FL017441

ACKNOWLEDGEMENTS

The Chinese Proficiency Test was developed through support provided by the Office of International Research and Studies of the U.S. Department of Education by John L.D. Clark of the Center for Applied Linguistics and Professor Ying-che Li of the University of Hawaii. The test Development Team included Professor Timothy Light of Ohio State University, Professor Albert Dien of the Chinese Language Teachers Association, and Professor Ronald Walton of the University of Pennsylvania.

TABLE OF CONTENTS

Overview of the CPT Program 1

History of the CPT 1
Test Content 1
Test Format 1

Uses of the CPT 2

Test Administration Procedures 2

Registering for the Test 2
Ordering the Test 3
Score Reporting and Billing 3

Interpretation of Test Scores 4

Statistical Characteristics of the Test 4

Reliability and Standard Error of Measurement 4
Validity 6
Item Analysis 8
Intercorrelations Among CPT Subscores 9

OVERVIEW OF THE CPT PROGRAM

History of the CPT

The Chinese Proficiency Test (CPT) was developed by the Foreign Language Education and Testing Division of the Center for Applied Linguistics (CAL) in 1983. It was made possible by a grant from the U.S. Department of Education, with the collaboration of Chinese language scholars and under the general supervision of national advisory and test development committees of Chinese language instructors and program coordinators.

Preliminary norming administrations of the CPT were conducted in the fall of 1983 at 41 colleges and universities in the United States and abroad. A total of 617 students participated in the initial norming administrations. In the subsequent four years, over 700 students from 80 institutions took the CPT.

Test Content

The CPT is designed to evaluate the proficiency in Chinese listening and reading comprehension attained by Americans and other English-speaking learners of Chinese. In keeping with its intended purpose as a measure of general language proficiency, the major focus of the test is on assessing the student's ability to function in listening and reading situations typical of real-life language-use settings.

The level of difficulty of the test is appropriate to differentiate among students who have completed at least one year of regular college-level instruction in Chinese or an equivalent number of contact hours in high school. The range of the test content is sufficiently broad to permit its use with students at all subsequent levels of study, up to undergraduate or graduate majors in Chinese at the completion of their language programs.

Test Format

The CPT consists of 150 4-option multiple-choice items. All responses are gridded on separate, machine-scored answer sheets. The actual testing time is approximately two hours; however, about three hours are needed for the entire test administration.

The CPT is divided into two sections: Listening Comprehension (60 items) and Reading Comprehension (90 items). In the Listening Comprehension section, the student listens to a number of spoken dialogues between two or more speakers, and longer single-speaker passages such as news broadcasts, public interest announcements, and so forth recorded on a cassette tape. Following each passage, the student is asked one or more questions for which he or she must choose the correct answer from among four options printed in the test booklet. All Chinese voices on the tape are of native Mandarin Chinese speakers, carefully selected for voice quality and authenticity of expression. The spoken material is delivered with natural pacing and intonation. The response options are printed in English in the test booklets.

The Reading Comprehension section has two subsections: Structure (35 items) and Reading (55 items). The Structure subsection assesses the student's ability to

recognize correct structural patterns in written Chinese. The Reading subsection measures the student's comprehension of a variety of printed texts ranging in length from a few characters (e.g., street or store-front signs, newspaper headings) to passages of paragraph length (e.g., excerpts from newspaper or magazine articles, personal and official correspondence, short stories). All test questions deal with either the factual content of the passage or with straightforward inferences based on the passage. There are no texts or questions requiring a background in Chinese literature.

In order to make the test equally appropriate for students whose exposure to written Chinese has been based on either traditional or simplified characters, all stimulus material in both subsections of the Reading Comprehension section is printed, in parallel form, in both traditional and simplified characters. (For sample test questions, see the CPT Examinee Handbook.)

USES OF THE CPT

The primary focus of the CPT is on assessing the student's ability to understand spoken and written Chinese encountered in contemporary, real-life language-use contexts. As such, a major use of the CPT is to provide an objective measure of an individual's general proficiency in Chinese. Another major use of the CPT is to provide Chinese-teaching institutions with an external measure of the overall effectiveness of their instructional programs. A number of other uses are possible as well. The CPT has been used for the following purposes:

* admission to a Chinese study program
* placement within a Chinese study program
* exemption from a Chinese language requirement
* application for scholarship or appointment
* competency testing upon exit from a Chinese program
* certification of Chinese language proficiency for career placement
* evaluation of a Chinese instructional program

The CPT program encourages use of the test by an institution if such use will assist it in making valid decisions in one or more of the above applications. However, the institution itself must determine whether the CPT is an appropriate test with respect to its intended uses, and must establish its own level of acceptable performance on the test.

CPT score users are invited to consult with CPT program staff about current or intended uses of test scores. The CPT office can assist institutions contemplating use of the test by providing information about the applicability of the test to particular situations and by providing assistance in establishing performance criteria on a local basis.

TEST ADMINISTRATION PROCEDURES

Registering for the Test

The CPT is available to colleges, universities, and other institutions in the United States and abroad for administration to their students on dates selected by

the institution. An individual wishing to take the CPT first contacts CAL for the name of the nearest CPT test center. If a test center is not available nearby, the individual should contact a local institution that offers a Chinese language program. The institution may then apply to CAL to administer the CPT by designating a supervisor and filling out an order form (see "Ordering the Test" below).

When registering for the CPT, an examinee is required to present a photo identification document (e.g., driver's license, student ID, or passport) to the supervisor. The supervisor must verify the identification and record the name and social security number. The $20 testing fee is payable either at the time of registration or on the test date. No examinee should be allowed to take the CPT without paying the fee in advance. The testing fee can be paid in cash or by a check payable to the test center supervisor or to the institution. An invoice for the total testing fee will be issued to the institution after the test has been administered and test materials returned to CAL.

Ordering the Test

To place an order for the CPT, the supervisor must provide CAL with the following information in writing:

1. title/position and affiliation of the supervisor;
2. approximate test date;
3. number of examinees expected;
4. whether rush delivery is desired (at the institution's expense).

A CPT test order should arrive at the CPT Program Office at CAL approximately 2-3 weeks before the intended test date to allow timely shipment of the necessary test materials. To expedite processing of test orders, it is highly recommended that printed CPT Order Forms be used. (A copy is printed on the last page of this manual.) Telephone orders are accepted only under urgent circumstances and only from institutions which have administered the CPT previously.

Upon receipt of the Order Form, CAL will forward to the designated supervisor the requested number of CPT test booklets and associated materials, including the test tape for the Listening Comprehension section, student answer sheets, examinee handbooks, the supervisor's manual, and other documents.

It is the responsibility of the supervisor to safeguard and maintain the security of the test materials from the time they are received at the institution until they are returned to CAL. It is important that the test be administered in strict accordance with the procedures outlined in the supervisor's manual. All materials requested on the Test Materials Checklist must be accounted for and returned to CAL (see CPT Supervisor's Manual for detailed instructions).

Score Reporting and Billing

Approximately 2-3 weeks following receipt of the returned test materials, CAL will forward a roster of scores to the institution. An invoice for the total testing fee, including any delivery expense for rush orders, will follow shortly afterward. This invoice must be paid by check or money order, payable to: Chinese Proficiency Test, Center for Applied Linguistics.

INTERPRETATION OF TEST SCORES

The CPT score roster lists the total score and the following three subscores for each student tested: Listening Comprehension, Structure, and Reading. All are raw scores based on the total number of correct responses on each subsection of the test. The maximum possible raw score for each part is: 60 for Listening Comprehension, 35 for Structure, and 55 for Reading. No penalty for guessing is imposed for incorrect responses, and blanks and non-responses are counted as incorrect answers.

The CPT is a norm-referenced test. This means that CPT scores can be compared with those of others in the United States and abroad who have taken the test (called the norming population). The current norming population consists of 479 students who took the CPT between 1984 and 1987 at some 60 institutions in the United States and abroad. At the time of testing, these students completed a background questionnaire containing the question: "What is the highest level of Chinese language course (not literature) you are PRESENTLY ENROLLED IN?" Response options to this question were: beginning, intermediate, and advanced. On the basis of student responses to this question, three separate norming groups--beginning, intermediate, and advanced--were identified from the total population. CPT staff assume that beginning students were enrolled in a first-year, college-level Chinese course, and that intermediate students were enrolled in a second-year course. The total number of students in the three groups was 112, 220, and 147, respectively.

To compare a student's performance with the performance of the norming population, refer to the CPT Percentile Rank Table on the following page. The CPT Percentile Rank Table contains percentile rank data on the three groups of students. To use this table, first identify the student as beginning, intermediate, or advanced. Then, look up the raw scores in the left-most column of the table, and read across to find the corresponding percentile ranks. For example, an intermediate student who obtained a Listening raw score of 28 would rank 55% in the norming group, meaning that about 55% of the intermediate students in the nation scored lower than this student on the Listening Comprehension section.

In addition to making use of the normative data provided in this manual, institutions may wish to validate the CPT for a particular test use or to carry out further analyses of CPT scores on a local basis. CPT staff would be pleased to discuss such analyses or validation studies with interested institutions.

STATISTICAL CHARACTERISTICS OF THE TEST

Reliability and Standard Error of Measurement

The reliability of a test is the extent to which it yields consistent results. The standard error of measurement (SEM) is an estimate of the probable extent of the error inherent in a test score due to the imperfect reliability of the test. Thus, the higher the reliability, the lower the SEM. Together, these two indices provide an indication of how consistently a test measures student performance.

Table 1. CPT Percentile Rank Table.

Raw Score	Beginning			Intermediate			Advanced		
	List.	Str.	Read.	List.	Str.	Read.	List.	Str.	Read.
60									
59									
58									
57									
56									
55									
54									
53							99		
52				99					99
51							97		97
50				98					95
49				97			97		94
48				97			93		92
47				96			90		
46						99	88		89
45				96			86		87
44				96		98	83		84
43				95		98	80		81
42	99					97	79		80
41				93		96	77		77
40	97		99	91		95	71		76
39				91		94	68		74
38	95			90		91	64		69
37				87		90	62		65
36	92			86			57		61
35	91		98	82		89	52		59
34	89			78		88	47		56
33				72		87	42	99	53
32	86			70		85	39	98	51
31	84		97	67		83	32	97	48
30	81			63		78	29	95	42
29	78		96	58		76	25	94	37
28	77		95	55	99	71	21	92	32
27	76		93	52		69	17	91	27
26	70		90	48	98	63	16	88	24
25	68		89	43	98	56	15	84	21
24	65		87	39	96	51	12	82	18
23	60		80	36	94	45	10	79	16
22	55		73	33	91	41	6	71	15
21	51	99	63	28	88	37	5	69	13
20	43	97	62	23	86	32	4	59	10
19	34	96	55	20	81	30	3	55	8
18	30	94	55	16	77	24		50	6
17	24	91	53	11	69	21		42	5
16	20	88	45	8	63	19	2	33	3
15	13	85	42	6	56	13	1	26	
14	9	77	38	4	48	11		22	
13	5	69	34	3	42	8		16	3
12	4	61	32	3	34	6	1	12	
11	3	53	30	1	30	6		8	2
10	1	40	29	1	21	5		6	
9		30			14	4		5	
8		25	28		11	3		3	
7		13	27		7	3		1	1
6		7	26		5				
5		3			3	2			
4		2			2				
3					1				
2		1			1				
1									
0			25[a]			1			

Note. This data is based on the performance of 479 examinees who took the CPT between 1984-1987.

[a] These students did not respond to any question in the Reading section, probably because of its relative difficulty.

The table below gives the Kuder-Richardson (Formula 20) reliabilities and SEMs of CPT scores based on the 1984-1987 data.

Table 2. Reliabilities and Standard Errors of Measurements of CPT Scores.

Section	Reliability	SEM
1. Listening Comprehension	.89	3.32
2. Structure	.83	2.61
3. Reading	.93	3.22
TOTAL	.96	5.14

The SEM is very useful in interpreting a test score. Because every test involves some amount of measurement error, an observed test score is made up of a "true score" and a "measurement error." That is,

Observed score = true score + measurement error.

It is assumed that the observed scores are normally distributed about the true score. The standard deviation of that distribution is the standard error of measurement. Therefore, for Listening Comprehension, approximately 95% of the examinees have their true scores within 6.64 (3.32 x 2) points of their reported scores. Alternatively, for an obtained Listening Comprehension score of 30, we can be 95% sure that the true score falls in the range of 23.36 (30 - 6.64) and 36.64 (30 + 6.64).

A related important use of the SEM is to determine whether an observed score difference between two students reflects "true" difference in performance above and beyond variations attributable to measurement error. Two scores are considered to represent significantly different true scores only when they differ by more than twice the SEM. For example, scores obtained by two students on the Listening Comprehension section should not be considered to represent significant differences unless they are at least 6.64 points (3.32 x 2) apart. Based on this guideline, the minimum values for two scores to be considered as representing different true scores on each subtest and the total test are as follows: 6.64 for Listening Comprehension, 5.22 for Structure, 6.44 for Reading, and 10.28 for the total test.

Validity

Validity refers to the extent to which a test measures what it is intended to measure. For the CPT, the question of validity refers to how well the test measures a person's proficiency in the Chinese language. Although a test publisher may present qualitative and empirical evidence of validity, it is ultimately the test user who must determine whether a particular test is valid for a particular purpose and for a particular examinee. Among the many types of test validity, two, content validity and concurrent validity, are discussed here.

Content validity refers to the extent to which the items in a test adequately sample the linguistic and behavioral domains of the construct (i.e., Chinese proficiency) that the test is meant to measure. The degree to which a test is content valid is determined by examining the degree of similarity between the stimulus passages and the corresponding language-use situations the examinee might be expected to encounter in real-life language use.

In constructing CPT items, certain content domains were postulated as appropriate to the proficiency of an English-speaking learner of Chinese. Representative items were then developed to reflect each domain. The content domains that were identified can be grouped into three categories, representing increasing levels of proficiency in Chinese as a second language. These levels are:

Level One--Survival. At this level, the learner can "survive" when confronting simple daily tasks such as shopping for food, clothes, and incidental necessities of life. CPT listening comprehension items at this level require the understanding of oral language heard in a grocery story, train station, bank, bus, airport, etc. CPT reading comprehension items at this level include certain street signs, storefront designations, and other similar "sight" vocabulary encountered in daily situations.

Level Two--Tourist Resident in Chinese Society. At this level, the learner can understand social conversation on everyday topics and can read short descriptive narrations. CPT listening comprehension items at this level require comprehension of short conversations between two people involving a single conversational turn. The conversations deal with such topics as planned trips, family problems, accidents, etc. CPT reading comprehension items at level two require the examinee to comprehend newspaper headlines, printed announcements, and simple descriptions from newspapers and magazines.

Level Three--Foreign Professional in Chinese Society. At this level, the learner can comprehend oral and written language on virtually any matter of general interest to educated native speakers in authentic Chinese situations. CPT listening stimuli at this level involve two conversational turns and may deal with topics of a general academic nature such as recent advances in health care, current events, etc. Questions may deal with the speakers' opinion or attitude, or they may require inference on the part of the examinee. Printed stimuli at this level are paragraphs followed by one or more questions dealing with basic factual information covered in the passage, or inference or analogy based on information presented in the passage. The passages cover a wide variety of daily topics such as those that may be found in a newspaper or magazine (e.g., results of a public opinion survey, job descriptions, vacation travel, etc.).

Items testing recognitional knowledge of Chinese syntax are of two types. The first type tests word order by requiring the examinee to choose, from among four possible positions within a sentence, the correct position for a word or phrase in Chinese. The second type of item tests knowledge of correct structure by requiring the examinee to choose from among printed four options the correct word or phrase that should be inserted in a blank within a printed sentence.

Validity may also be examined by determining the extent to which a test score corroborates the results that may be obtained through use of an independent external criterion at the same point in time. This approach may be referred to as evidence

of concurrent validity with another measure. In this manual, self-reported proficiency level is used as the external criterion for assessing concurrent validity.

Table 3 reports the descriptive statistics of the CPT scores obtained by examinees at three levels of proficiency--beginning, intermediate, and advanced (see page 4 for categorization of the three proficiency levels). High concurrent validity of the CPT is supported by the observed substantially higher mean scores of the "Advanced" students in comparison to the "Intermediate" students. Similarly, higher mean scores of the Intermediate students are also observed in comparison to those of the Beginning students. Thus, the CPT score patterns appear to corroborate and reflect self-reported proficiency levels indicated by the examinee at the time of testing.

Table 3. Means and Standard Deviations of CPT Subscores.

	Beginning			Intermediate			Advanced		
Section	N	Mean	SD	N	Mean	SD	N	Mean	SD
1. Listening	109	23.33	7.70	220	27.82	9.05	145	35.30	8.70
2. Structure	112	11.77	4.19	220	14.91	5.38	147	19.22	5.91
3. Reading	112	15.35	10.66	220	24.51	8.92	147	33.35	9.99
TOTAL	109	50.13	16.98	220	67.24	20.03	145	87.97	22.64

Item Analysis

The level of difficulty (sometimes called "facility index" or "item difficulty") is measured as the percentage of correct responses for each item. It indicates how difficult an item is for a group of examinees. As such, the level of difficulty of a test (average of item difficulty levels) depends both on the inherent difficulty of the test and on the ability level of the examinees who take the test. Both factors determine whether test difficulty is appropriate for the group of examinees.

For a test to be of middle difficulty, the average raw score of a test should be midway between a perfect score and the score that would be expected if the answer to each question were marked at random. For the CPT, this optimal value is 62.5% of the total number of questions in a section. The difficulty indices of the CPT subsections and the total test are given in Table 4.

Table 4. Difficulty indices of CPT Subsections.

Section	Difficulty index
1. Listening comprehension	.50
2. Structure	.45
3. Reading	.47
TOTAL	.48

Table 4 shows that the CPT is a relatively difficult test, with the three subsections having a difficulty index of .50 for Listening, .45 for Structure, and .47 for Reading, all below the optimal 62.5% level. This is perhaps due to the fact that the CPT was designed primarily for intermediate and advanced students of Chinese, but approximately one quarter of the norming population (112 out of 479) were beginning students. For intermediate- and advanced-level students, the CPT is a better discriminator of differences in proficiency than it is for beginning-level students.

Interrcorrelations Among CPT Subscores

The three subsections of the CPT are designed to measure different skills within the general domain of Chinese proficiency. It is expected that these skills are interrelated; i.e., persons who are highly proficient in one skill area tend to be proficient in the other areas as well. However, the intercorrelations are not expected to be perfect. If they were, there would be no need to report scores for each section, and the subscores would represent the same rather than different aspects of language proficiency.

Table 5 reports the Pearson Product-Moment correlation coefficients measuring the extent of relationships among the three subsections and the total score based on the 1984-1987 data. A correlation coefficient of 1.0 would indicate a perfect relationship between two sections, and 0.0 would indicate a total lack of relationship. The table shows that there is a fairly strong relationship among the skills tested by the three subsections of the CPT, but each of the subsection scores provides some unique information about the examinee's proficiency in the Chinese language. Note that the most unique information seems to be provided by the Reading subscore, since Reading shows less relationship to Listening (.69) and Structure (.71) than Structure shows to Listening (.78). Thus, Structure shows greater overlap between the two skills (Listening and Reading) than either skill overlaps with the other.

Table 5. Intercorrelations Among the CPT Subscores.

Section	1	2	3
1. Listening Comprehension	-		
2. Structure	.78	-	
3. Reading	.69	.71	-
TOTAL	.91	.88	.91

CHINESE PROFICIENCY TEST

Order Form

Ship to: (Please print)

Name

Position Title

Department

Institution

Street

City/State/Zip Code

Telephone #

Bill to: (if different from left)

Name

Position Title

Department

Institution

Street

City/State/Zip Code

Telephone #

First order?

____ Yes (If yes, CAL reserves the right to verify the above information)
____ No

Number of examinees expected: _____ ($20 per examinee)

Purpose and use of the test: (Check all that apply)

____ admission to a Chinese study program
____ placement within a Chinese study program
____ exemption from Chinese language requirement
____ application for scholarship or appointment
____ competency testing upon exit from a Chinese program
____ certification of Chinese language proficiency for career purposes
____ other (Please specify: _____)

Approximate testing date: _____ **Required delivery date:** _____

Payment:

You will be billed for the actual number of students tested after you return the test materials to CAL. Delivery charges will be paid by CAL, provided this Order Form reaches CAL at least two weeks prior to the required delivery date indicated above. Delivery charges for rush orders requiring a shorter delivery time will be billed to the institution placing the order. Official score reports will be issued about two weeks after the materials have been returned to CAL.

Security Agreement

It is vital that you maintain the security of the test materials from the time you receive them until they are returned to CAL. You must agree to: (1) keep the CPT test booklet(s) and test tape(s) confidential and secure, and (2) guarantee that none of the test booklet(s) or test tapes(s) will be reproduced in any form. Please signify your acceptance of these terms by signing your full name below.

Signature Date

Office Use Only

Date received: _____ Date filled: _____ Processed by: _____

Detach here

The Center for Applied Linguistics (CAL) was chartered in 1959 as a non-profit corporation in Washington, DC, with initial support provided by the Ford Foundation. While historically CAL has developed language tests to meet the needs of its various domestic and international projects, in 1986 CAL's language testing initiatives were centralized in the Foreign Language Education and Testing Division. The Division produces second language tests for use in governmental, educational, and occupational settings. For information on CAL's test development capabilities, contact Charles W. Stansfield, Director, Division of Foreign Language Education and Testing, CAL, 1118 22nd Street, NW, Washington, DC 20037, (202) 429-9292.